

Figure 1: Top 10 BLASTP results for INTP039 against NCBI non-redundant database.

		Score (bits)	E Value
gb AAH39632.1	Similar to protease, serine, 8 (prostasin) [Mus musculus]	714	0.0
dbj BAC11431.1	Unnamed protein product [Homo sapiens]	266	6e-70
ref XP_110113.1	RIKEN cDNA 1300008A22 [Mus musculus]	173	7e-42
ref NP_114154.1	Marapsin; channel-activating protease 2 [Homo sapiens]	172	9e-42
gb AAH36846.1	Similar to protease, serine, 8 (Prostasin) [Homo sapiens]	172	1e-41
ref NP_690851.1	Serine protease EOS [Homo sapiens]	172	2e-41
emb CAC85953.1	Matriptase-2 [Homo sapiens]	167	3e-40
ref NP_705837.1	Type II transmembrane serine protease 6 [Homo sapiens]	167	3e-40
ref XP_128466.1	Similar to RIKEN cDNA [Mus musculus]	165	2e-39
gb AAG32641.1	Prostasin [Rattus norvegicus]	165	2e-39

Figure 2: Alignment between INTP039 polypeptide sequence (SEQ ID NO:22) and the hit, NP_114154.1.

```

ref|NP_114154.1| marapsin; channel-activating protease 2 [Homo sapiens]
sp|Q9BQR3|MPN_HUMAN Marapsin precursor
emb|CAC35467.1| marapsin [Homo sapiens]
gb|AAK38168.1| pancreasin [Homo sapiens]
      Length = 290

Score = 172 bits (437), Expect = 9e-42
Identities = 88/236 (37%), Positives = 132/236 (55%), Gaps = 12/236 (5%)

Query: 42  NTVPGGEWPWQASVRRQGAHICSGSLVADTWVLTAAHCFEKAATELNSWSVVLGSLQREG 101
      +T GEWPWQ S++R G+H C GSL+A+ WVLTAAHCF + T L + V+LG+ Q
Sbjct: 40  DTQEGEWPWQVSIQRNGSHFCGGSLIAEQWVLTAAHCFRNTSETSL--YQVLLGARQLVQ 97

Query: 102 LSPGAEEVGVAALQLPRAYNHYSQGSDDLALLQLAHPTTHT----PLCLPQPAHRFPFGAS 157
      P A V ++ Y + +D+AL++L P T P+CLP P+ F G +
Sbjct: 98  PGPHAMYARVRQVESNPLYQGTASSADVALVELEAPVPFTNYILPVCLPDPSVIFETGMN 157

Query: 158 CWATGW----DQDTS DAPGTLRNLRLRLISRPTCNCIYNQ-LHQRHLSNPARPGMLCGGP 212
      CW TGW ++D P L+ L + +I P CN +Y++ + + MLC G
Sbjct: 158 CWVTGWGSPSEEDLLPEPRILQKLAVPIIDTPKCNLLYSKDTEFGYQPKTIKNDMLCAGF 217

Query: 213 QPGVQGPCQGDSSGPVLCLEPDGHVWQAGIISFASSCAQEDAPVLLTNTAAHSSWL 268
      + G + C+GDSGGP++CL W+QAG+IS+ CA+++ P + AH +W+
Sbjct: 218 EEGKKDACKGDSGGPLVCLVGQS-WLQAGVISWGEGCARQNRPGVYIRVTAHHNWI 272

Score = 104 bits (259), Expect = 4e-21
Identities = 75/251 (29%), Positives = 110/251 (42%), Gaps = 21/251 (8%)

Query: 292 SCVACGSLRT----AGPQAGAPSPWPWEARLMHQQLACGGALVSEEAVLTAAHCFIGRQ 347
      + ACG R G Q WPW+ + G CGG+L++E+ VLTAAHCF
Sbjct: 22  AATACGRPRMLNRMVGGQDTQEGEWPWQVSIQRNGSHFCGGSLIAEQWVLTAAHCFRNTS 81

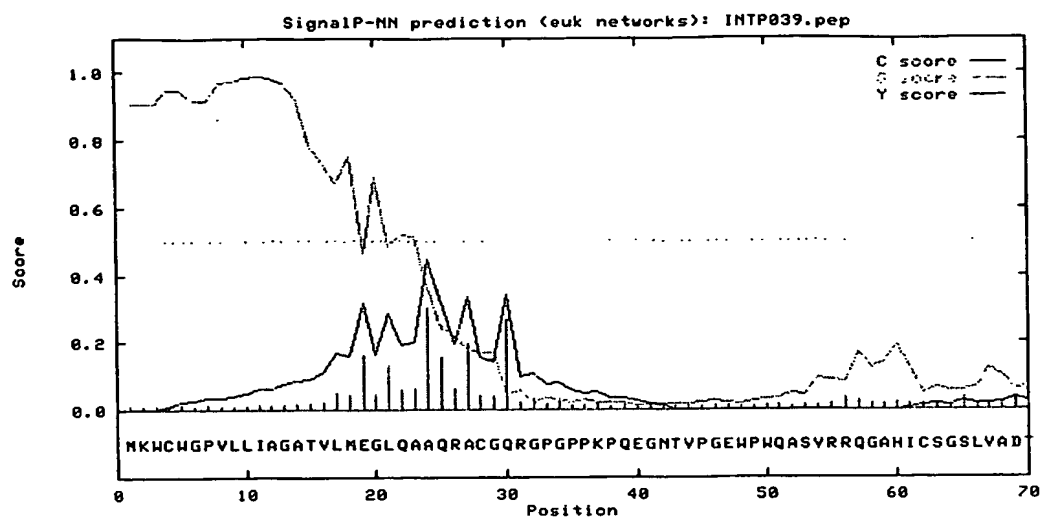
Query: 348 APEEWSVGLGTR-----PEEW--GLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLR 399
      + V LG R P ++Q+ + Y D+AL+ L PV +
Sbjct: 82  ETSLYQVLLGARQLVQPGPHAMYARVRQVESNPLYQGTASSADVALVELEAPVPFTNYIL 141

Query: 400 PLCLPYPDHHLPDGERGWVLGRARPG-----AGISSLQTVPVTLGPRACSRRLHAAP--- 451
      P+CLP P G WV G P LQ + V ++ C+ L++
Sbjct: 142 PVCLPDPSVIFETGMNCWVTGWGSPSEEDLLPEPRILQKLAVPIIDTPKCNLLYSKDTEF 201

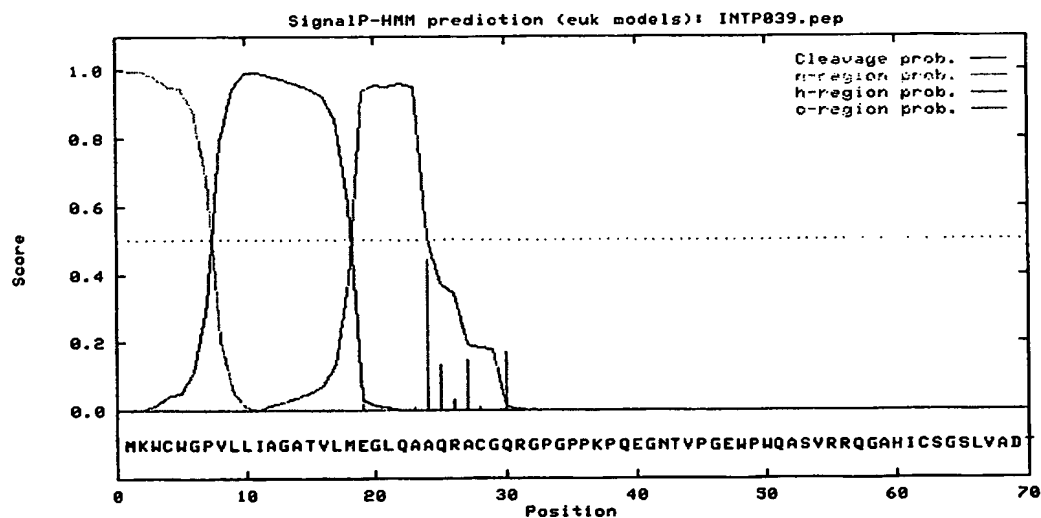
Query: 452 GGDGSPILPGMVCTS-AVGELPSCGSLGAPLVHEVRGTWFLAGLHSFGDACQGPAPPAV 510
      G I M+C G+ +C+G SG PLV V +W AG+ S+G+ C RP V
Sbjct: 202 GYQPKTIKNDMLCAGFEEGKKDACKGDSGGPLVCLVGQSWLQAGVISWGEGCARQNRPGV 261

Query: 511 FTALPAYEDWV 521
      + + A+ +W+
Sbjct: 262 YIRVTAHHNWI 272

```

Figure 3: SigP cleavage site prediction for INTP039.**SignalP-NN result:**

```
>INTP039.pep                      length = 70
# Measure  Position  Value  Cutoff  signal peptide?
max. C      24       0.306  0.33   NO
max. Y      24       0.447  0.32   YES
max. S      11       0.990  0.82   YES
mean S      1-23     0.820  0.47   YES
# Most likely cleavage site between pos. 23 and 24: LQA-AQ
```

SignalP-HMM**result:**

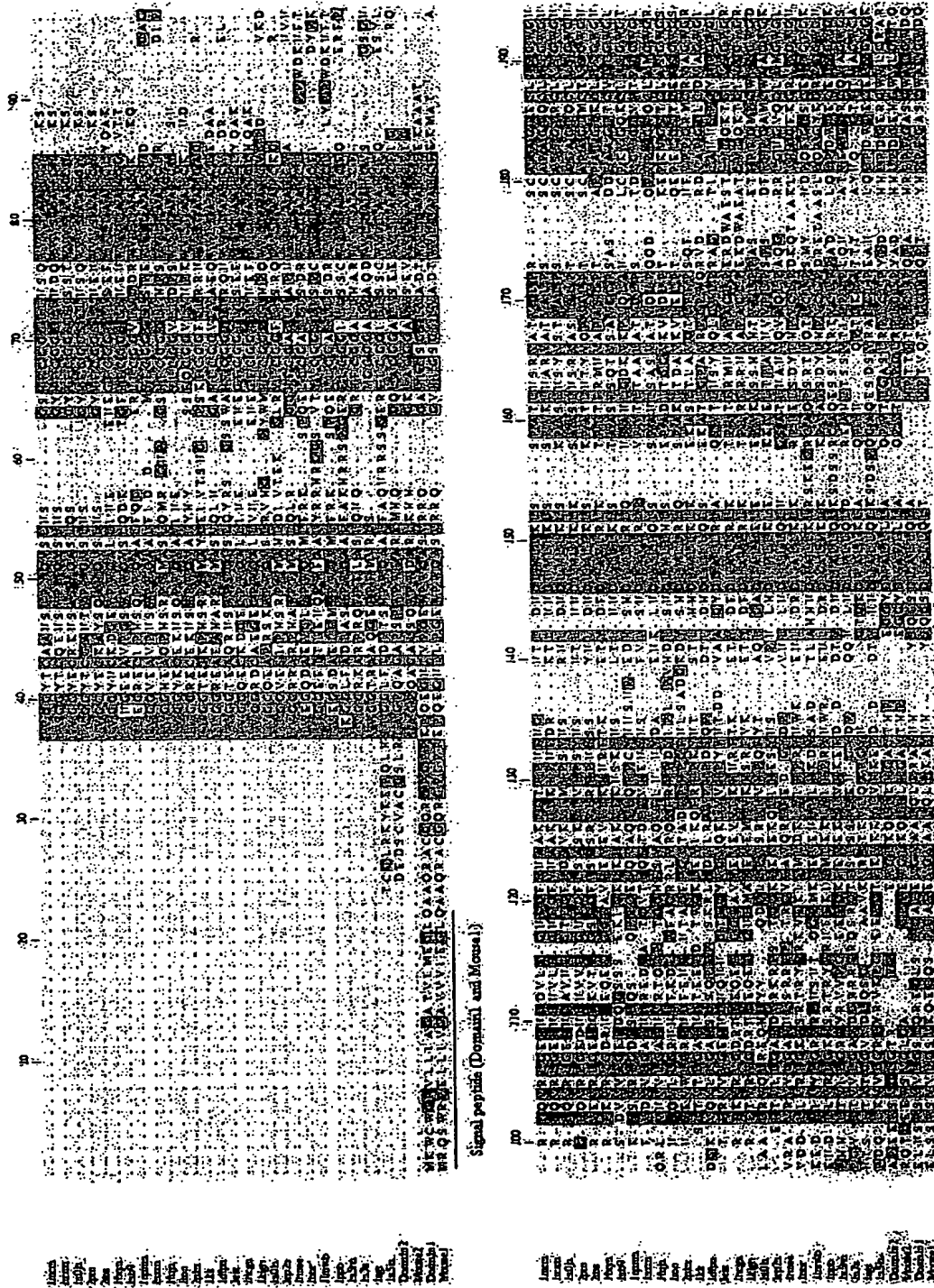
```
>INTP039.pep
Prediction: Signal peptide
Signal peptide probability: 0.998
Signal anchor probability: 0.000
Max cleavage site probability: 0.444 between pos. 23 and 24
```

Figure 4: INTP039 domain organisation

INTP039 of signal peptide followed by two Trypsin domains.



Figure 5: INTP039 multiple alignment



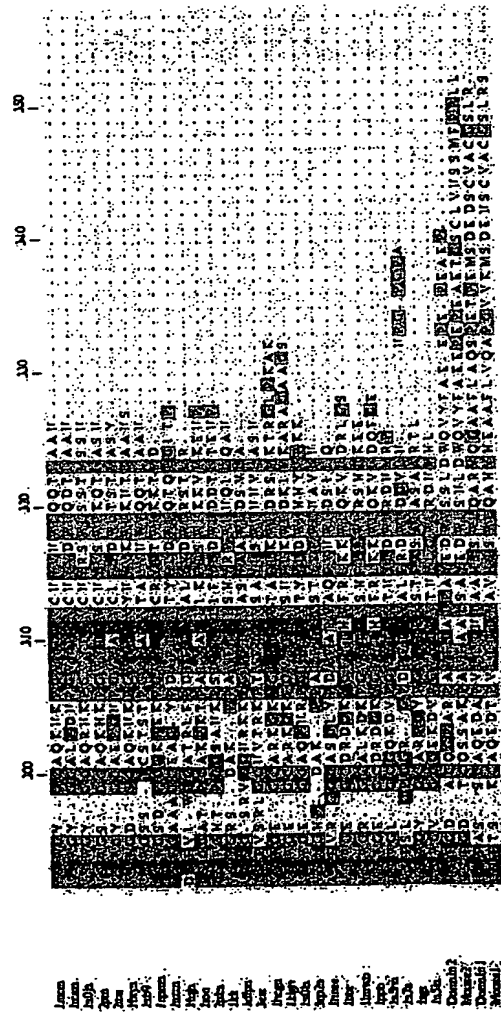
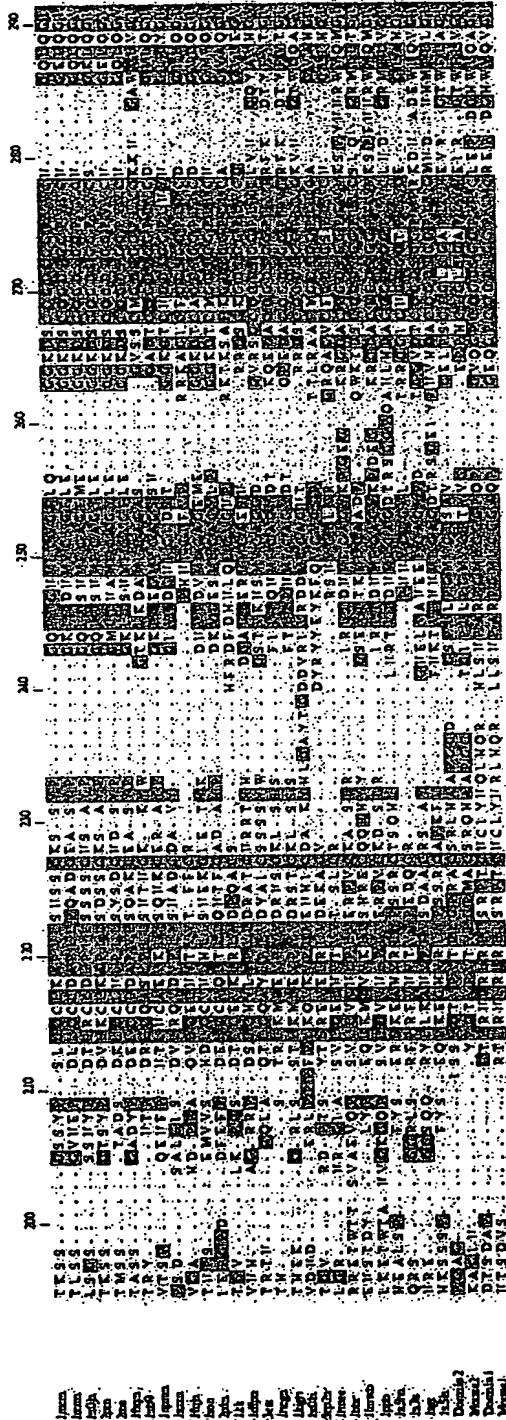
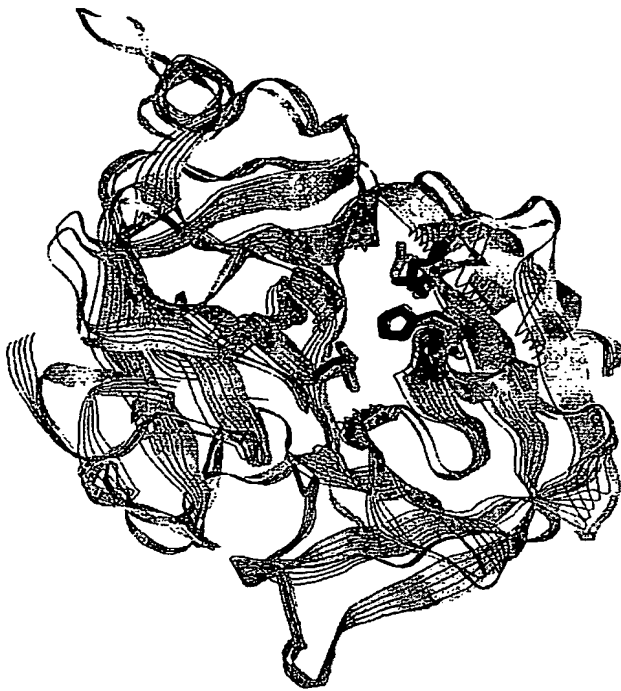


Figure 6: Model of INTP039 domains 1 and 2

Domain 1 (ribbon) : Template 1BUI (strand) Human Microplasmin 34% ID



Domain 2 (ribbon) : Template 1EAX (strand) Human Matriptase 33% ID

